

mirEX: a platform for comparative exploration of plant pri-miRNA expression data

Dawid Bielewicz¹, Jakub Dolata¹, Andrzej Zielezinski², Sylwia Alaba²,
Bogna Szarzynska¹, Michal W. Szczesniak², Artur Jarmolowski¹,
Zofia Szwejkowska-Kulinska^{1,2,*} and Wojciech M. Karlowski^{2,*}

¹Department of Gene Expression and ²Laboratory of Bioinformatics, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan, Poland

Received August 1, 2011; Revised and Accepted September 29, 2011

ABSTRACT

mirEX is a comprehensive platform for comparative analysis of primary microRNA expression data. RT-qPCR-based gene expression profiles are stored in a universal and expandable database scheme and wrapped by an intuitive user-friendly interface. A new way of accessing gene expression data in mirEX includes a simple mouse operated querying system and dynamic graphs for data mining analyses. In contrast to other publicly available databases, the mirEX interface allows a simultaneous comparison of expression levels between various microRNA genes in diverse organs and developmental stages. Currently, mirEX integrates information about the expression profile of 190 *Arabidopsis thaliana* pri-miRNAs in seven different developmental stages: seeds, seedlings and various organs of mature plants. Additionally, by providing RNA structural models, publicly available deep sequencing results, experimental procedure details and careful selection of auxiliary data in the form of web links, mirEX can function as a one-stop solution for *Arabidopsis* microRNA information. A web-based mirEX interface can be accessed at <http://bioinfo.amu.edu.pl/mirex>.

INTRODUCTION

MicroRNAs are the key post-transcriptional regulators of gene expression in Eukaryota. They control gene expression by targeting the cleavage of cognate mRNAs or by

inhibiting their translation (1–3). Therefore, when studying the biology of any organism, it is of utmost importance to quantify precisely the expression level of each particular microRNA gene during organ development. Northern hybridization, microarray analysis, deep sequencing approaches and real-time quantitative PCR (RT-qPCR) are standard techniques used to accomplish this task as described previously (4–7). The RT-qPCR is considered a gold standard method in precise quantification of gene transcript levels (8).

MicroRNA genes that encode transcripts which are processed to the same or similar mature microRNA species are grouped in families. In *Arabidopsis thaliana*, the number of such family members varies between 1 (e.g. miR163, miR173) and 14 (miR169) [miRBase release 17 (9)]. MicroRNA genes from the same family, although represented by identical or almost identical mature microRNAs, differ considerably in gene organization and sequence. In many cases it is only possible to observe the expression of all family members as a group, rather than the individual members due to the large sequence conservation when using northern hybridization or deep sequencing approaches. However, individual members of a given microRNA family may be expressed in different developmental stages or in response to various biotic/abiotic stimuli (10–12).

Our resource contains information regarding the expression of 190 *A. thaliana* microRNA genes at the level of primary microRNA (pri-miRNA) in different developmental stages obtained using a RT-qPCR high-throughput platform. Many databases that store and allow access to microRNA gene expression profiles from variety of organisms already exist [e.g. (13–18)], however,

*To whom correspondence should be addressed. Tel: +48 61 829 5841; Fax: +48 61 829 5949; Email: wmk@amu.edu.pl
Correspondence may also be addressed to Zofia Szwejkowska-Kulinska. Tel: +48 61 8295766; Fax: +48 61 829 5949; Email: zofszwey@amu.edu.pl
Present address:

Bogna Szarzynska, Institut de Biologie de l'Ecole Normale Supérieure, CNRS UMR8197, INSERM U1024, 46 rue d'Ulm, 75230 Paris Cedex 05, France.

The authors wish it to be known that, in their opinion, the first four authors should be regarded as joint First Authors.

they have various limitations and restrictions. One of the limitations includes the lack of tools and/or data for comparisons between developmental stages and various genes at the same time. With this in mind, we developed the mirEX platform as a comprehensive starting point for a comparative investigation of microRNA genes expression. Our database offers to the scientific community easily accessible data and is of interest to researchers working on the specific microRNA function, the expression profile of entire microRNA family members during a particular organ/developmental stage, or on microRNA biogenesis. Additionally, by creating the mirEX interface, we hope to propose a new database interface standard for comparative microRNA gene expression data mining.

WEB TOOLS AND DATA

Expression datasets

In the current release, the mirEX database includes expression data of microRNA genes from *A. thaliana*. To collect the data, we prepared a real-time PCR platform for 190 primary microRNA sequences. For each microRNA, a pair of specific primers amplifying a single product was designed; 10 microRNAs primer sequences were taken from Pant *et al.* (19). The qPCR reactions for one biological replicate included 12 different controls and standards, were carried out in parallel as described previously (20) using RNA template isolated from six developmental stages of *A. thaliana* Columbia-0 wild-type plants. RNA from seeds and siliques was isolated according to Ref. (21). The detailed procedure of running the platform can be found in the mirEX database documentation.

The plant material used in expression profiling experiments includes: seeds, 10-day-old seedlings, 14-day-old seedlings, 25-day-old plants, 35-day-old plants, 42-day-old plants (rosette leaves and stems) and 53-day-old plants (rosette leaves, stems, inflorescence and siliques). For better visualization of the individual developmental stages, we included pictures of actual plants used in our experiments. Our high-throughput platform contains an original and novel set of data on microRNA gene expression in dormant seeds. The PP2A (At1g13320) and actin (At3g18780) cDNAs were used as an expression reference (22). The mean measurements of three replicas were used to calculate the fold-change value and presented in a form of \log_{10} . In a case when the value of correlation coefficient of the three replicas was <0.995 , such data was labeled as 'low quality' and is not shown by default on the graphs, but indicated in gray in the data tables. Since the level of expression of most of the primary microRNAs is lower than the reference genes, the data shown on the graphs in mirEX is rescaled in order to avoid presenting it as a negative value. Rescaling shifts the zero value of the graph's y-axis to the basal expression level of the whole experiment. This allows showing the expression profile in a positive data range, but does not change the actual values.

Web interface

The mirEX database interface is designed to be used by a bench scientist on an everyday basis. Following a simplicity rule, the interface of mirEX has been built on only two types of result windows and a simple two-step querying system.

There are two ways to access data in mirEX: by searching for a particular microRNA or microRNA family, or by browsing the database content. A single input window allows searching for an individual or a family of microRNA by providing the numerical part of its identification (ID). When entered ID corresponds to a record of a single microRNA, the resulting page shows all of the details for this particular sequence (Figure 1). In the case of whole family ID, the search results are presented in the form of a line graph, showing expression data for all stages available in the database (Figure 2). The table located below the graph contains numerical values for data presented in the graph, including low quality measurements. The IDs shown in the table allow quick access to detailed information about particular microRNA.

Browsing the data in mirEX starts at the two-step selection process. The first step includes selection of developmental stages, which is followed by the selection of microRNAs. In mirEX, there are three ways to select microRNAs: (i) by typing their IDs in the input window, (ii) by uploading a file with a comma-separated list of microRNA IDs (that can be created in the first step) and (iii) by selecting individual or groups of sequences from the tree-like expandable menu. During the process of entering information using a keyboard, the mirEX interface will provide the list of available microRNAs. It is possible to mix all of the input methods—the resulting page will refer to all entered or selected IDs.

Presentation of search/browse results may differ depending on the number of selected developmental stages or microRNA genes (see 'Data mining' section). By default, the graph of expression profiles for various stages and microRNAs shows only high quality data. However, it is a user-defined option in mirEX whether the low quality measurements are presented on the graph. The expression values can be displayed using two reference genes: actin and PP2A.

The results for a single microRNA are presented in a record window in a form that is divided into distinct sections (Figure 1). The upper part of the record is dedicated to expression information presented in the form of a graph and two tables. When applicable, the graph contains information about the expression of other members of the microRNA family, that can be dynamically turned on and off. Additionally, specific shortcuts allow quick access to the records of any member of such family. The two tables, containing either reference-calculated or raw data, can be hidden to reduce the amount of information presented at any given moment on the single record window. Each graph or table in the mirEX database can be saved for later use by selecting an appropriate button.

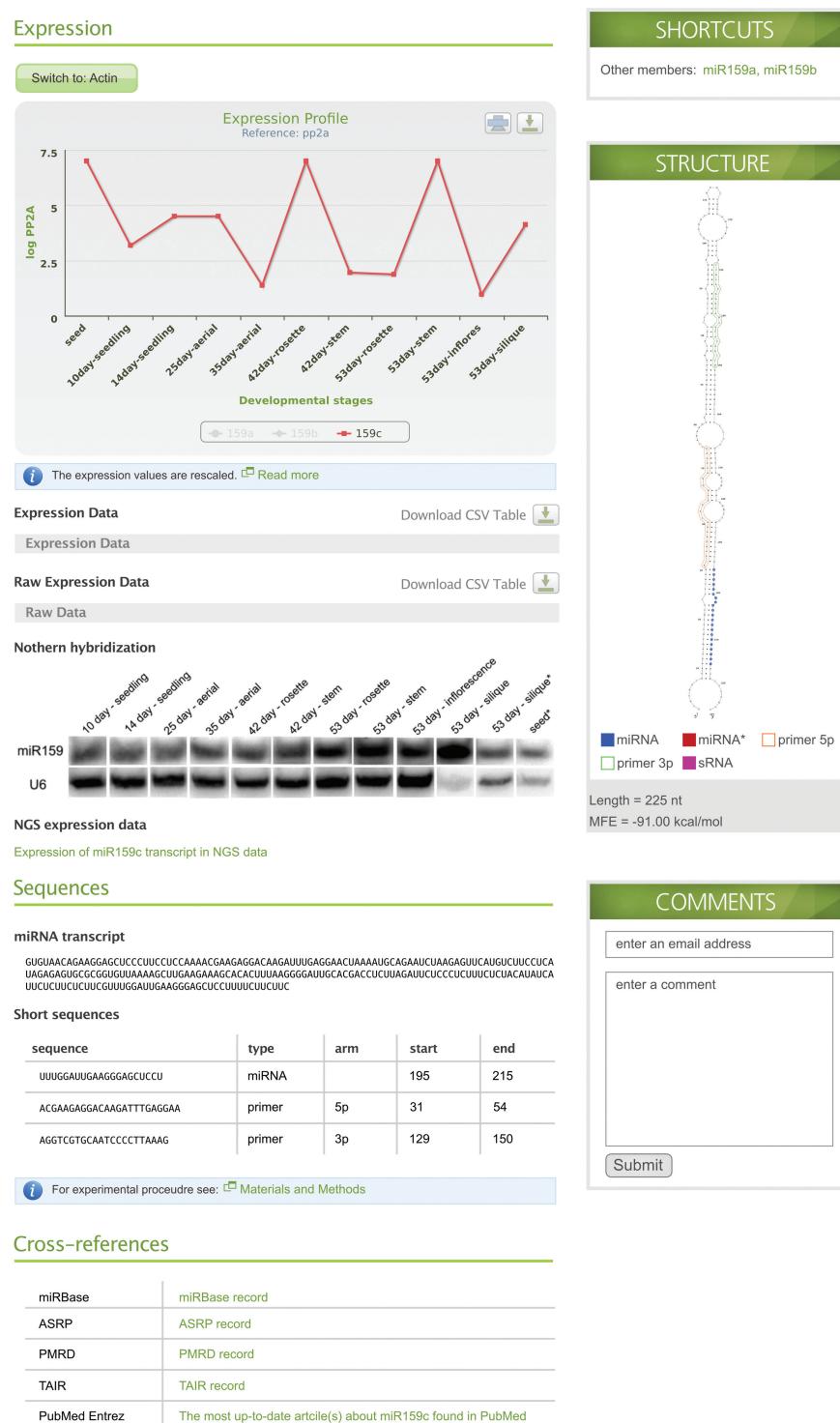


Figure 1. Example of mirEX record window for ath-miR159c. This window is divided into distinct sections representing: expression data, sequences and external references. The right panel contains shortcuts to microRNA family members' records, structure of the RNA transcript molecule with labeled mature sequences and RT-qPCR primers and webform for user comments.

Additionally, below the quantitative expression data, each record contains a web link to a text-based map of publicly available high-throughput Next-Generation Sequencing (NGS) short reads aligned to primary microRNA transcript. The information included here

contains sequence, location, length and number of the mapping reads. On a separate window, it is also possible to access the original record of the used NGS data.

The record window contains a graphical presentation of the structure of microRNA transcript with designations of

Expression

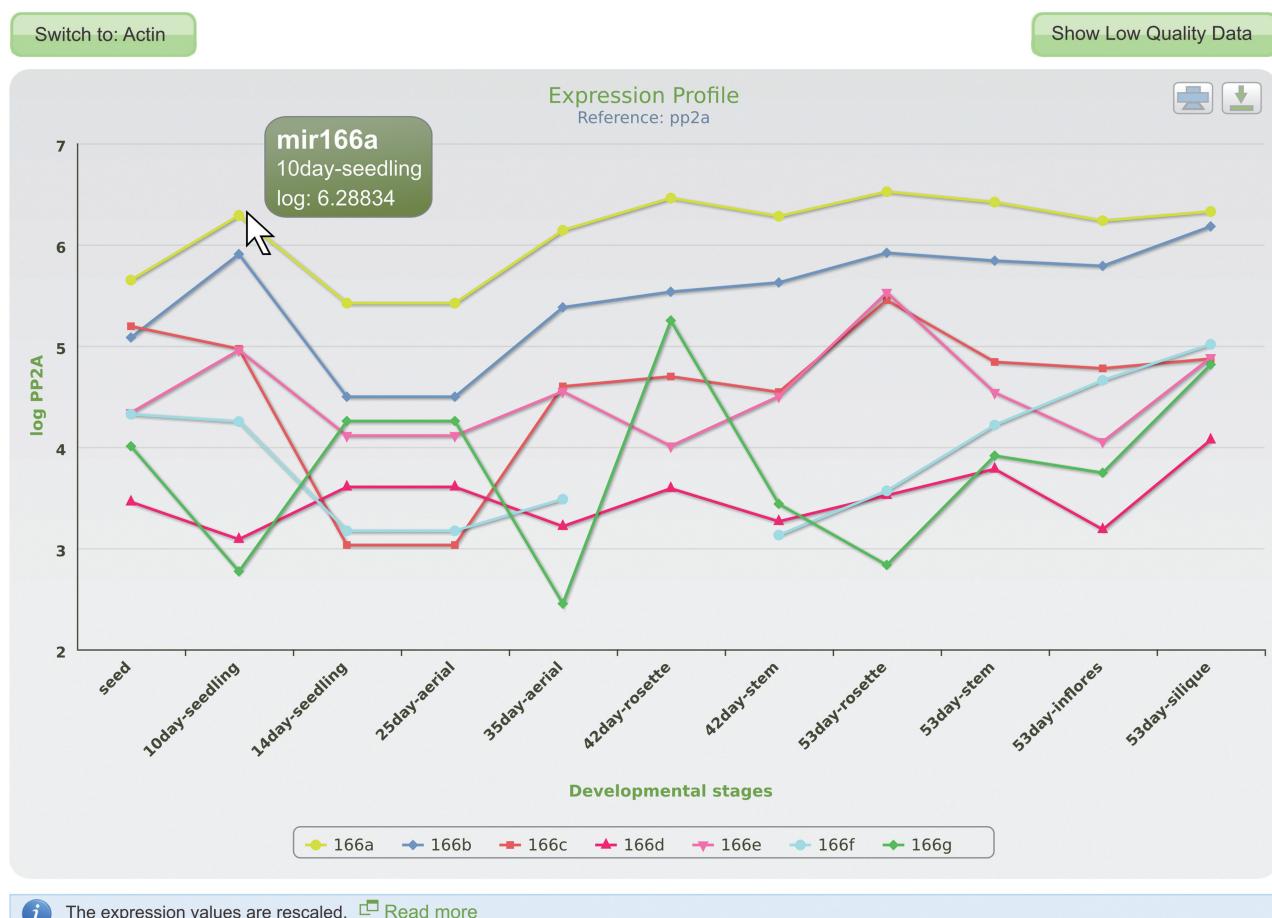


Table report

miRNA	seed	10day seedling	14day seedling	25day aerial	35day aerial	42day rosette	42day stem	53day rosette	53day stem	53day inflores	53day silique
166a	5.64921	6.28834	5.42083	5.42083	6.14107	6.45961	6.27966	6.52257	6.41905	6.23591	6.32645
166b	5.08157	5.90528	4.49621	4.49621	5.37973	5.53315	5.62505	5.91811	5.84033	5.78834	6.17987
166c	5.19281	4.96945	3.03096	3.03096	4.59848	4.69646	4.54260	5.45045	4.83994	4.77668	4.87147
166d	3.45695	3.08791	3.60479	3.60479	3.21583	3.59028	3.26447	3.52176	3.78184	3.18464	4.06962
166e	4.33894	4.95676	4.11181	4.11181	4.54865	4.00851	4.50100	5.52666	4.53720	4.05214	4.88188
166f	4.32301	4.25127	3.17132	3.17132	3.48388	4.57686	3.12957	3.56610	4.21639	4.65809	5.01434
166g	4.00704	2.77047	4.25605	4.25605	2.45153	5.24896	3.43708	2.83375	3.91363	3.74440	4.81495

i Low quality data are indicated in gray. Read more

Figure 2. Example of mirEX report window for ath-miR166 gene family. The report window contains graphical presentation of the expression levels, as well as actual numerical values presented in tabular format. Holding a mouse pointer over any data point allows access to details of expression measurements. The datasets for a particular microRNA can be dynamically turned on and off by clicking ID in the legend of the graph.

mature molecule(s) and primers used to assess its expression level by RT-qPCR. When available, the record data will be modified to include results from northern hybridization with mature microRNAs. However, it has to be noted that in most cases, such hybridization represents the

level of the whole microRNA gene family expression, and will be not specific for an individual microRNA gene. Moreover, the record window contains a table with sequences relevant to particular microRNA and a link to experimental procedures.

The last part of the record window includes references to external databases. Our aim was to avoid replication of the data available in other databases. For information such as gene structure or available papers, we direct the user to specialized databases, for example TAIR (23) or PubMed (24). The number of external references will grow in time, when new resources for microRNA biology are available, and in response to user requests.

To maintain a close relation with users of our database, we provide in the record window a simple tool for entering comments. The comment may be of any nature: concerning the biological aspect of the presented data or any general issue.

Although the interface is very intuitive and simple, at the main page we provide a video tutorial on all of the capabilities and various ways available to explore the mirEX tools and data. Moreover, every window in the mirEX interface contains context-specific help guides.

A complete 'road map' of the mirEX interface is included as [Supplementary Figure S1](#).

Data mining

The usefulness of the mirEX database lies in the ability to compare expression between different developmental stages and various microRNA genes. By using the 'browse' button, the user can, in two simple steps, select available developmental stages and a single or sets of microRNAs. Depending on the number of selected stages, the expression results are presented in the form of a bar graph (for one-stage analysis) or a line-based graph (for many stages). The bar graph presentation allows analysis of all pri-miRNAs at once with a zoom-in on an individual gene.

The line graph displays by default, data for only 24 microRNAs selected from the list. However, this restriction can be turned off. The line graph offers an additional option to dynamically remove and/or add any of the selected microRNA genes.

Another, quick way to access comparative expression profiles are shortcuts. In general, the shortcuts represent a new and easy way to explore data stored in the mirEX database. They are located in the same area on every page, and their content follows user selections. The use of shortcuts allows quick access to specific sets of data, e.g. the most expressed genes, or records representing microRNAs belonging to the same family. Periodically, we will modify the shortcuts in response to comments and readapt them according to the most frequently issued queries.

The strategy to use primary microRNA expression data allows fine exploration of gene activities of members of microRNA families. In some families, the expression pattern of individual genes in all developmental stages and organs is similar. However, the expression level of the individual pri-miRNAs may differ considerably: the fold change in the primary transcript levels within one microRNA gene family may range between 10 and 10⁵ (e.g. miR160, miR162, miR164, miR165, miR167, miR168, miR394, miR396, miR398, miR447 families, [Figure 3A](#)). Moreover, a few individual pri-miRNAs

show profound differences in their expression depending on the developmental stage and/or organ studied (e.g. miR156, miR157, miR158, miR159, miR166, miR169, miR397, miR399, miR404, [Figure 3B](#)). These differences may reflect either the existence of promoter regulatory elements that are responsive to developmental stimuli, or the various rate of pri-miRNA maturation during plant growth and organ formation. MicroRNA families containing a single representative also may exhibit dramatic changes in their expression levels (reaching even 10⁶-fold change) during plant growth and organ development (e.g. miR173, miR774, miR776, miR778, miR780, miR783, [Figure 3C](#)). Conversely, there are also pri-miRNAs that show relatively small fluctuations of expression levels during plant growth and organ formation. In conclusion, our data shows that each microRNA gene has its own characteristic expression profile reflecting its spatial and temporal regulation that can be followed and comparatively analyzed using tools implemented in the mirEX platform.

CONCLUSIONS

By creating the mirEX platform, we provide the scientific community with a high quality pri-miRNA expression data in seven developmental stages represented by 11 distinct organs of *A. thaliana*. This is a new and user-friendly platform designed to explore expression data in various developmental stages for a large number of related genes. The querying system has been limited to two simple steps, which allow access to any type of data stored in mirEX. Additionally, the mirEX interface does not require the use of a keyboard—the database user interface is fully mouse operated. This makes the process of selection and comparing data very easy and effective. Moreover, graphs presenting expression data are designed to accommodate user selection dynamically, which makes exploration of the mirEX content even more efficient.

The data currently incorporated in mirEX represents the starting point in our database development. The modular character of the database design makes it possible for further mirEX expansion to incorporate new species and datasets. Following, we plan to include *Arabidopsis* microRNA genes discovered in the future, various mutants and microRNA expression profiles from other plant species. The work on barley microRNA expression profiling is already underway. Incorporation of data from other species will broaden the available tool set allowing comparative analyses within and between species.

By designing the mirEX interface, we would like to propose a new trend in biological databases for simplicity and user-friendliness. Additionally, we put special attention to the web browser compatibility issue of the interface by testing all of the most popular tools. Careful selection of informatics techniques resulted in the platform that can be accessed even via iOS on mobile devices without loosing any of the functionality and interface features.

Carefully selected links to external databases, prevent the user interface from overloading with data, yet creates

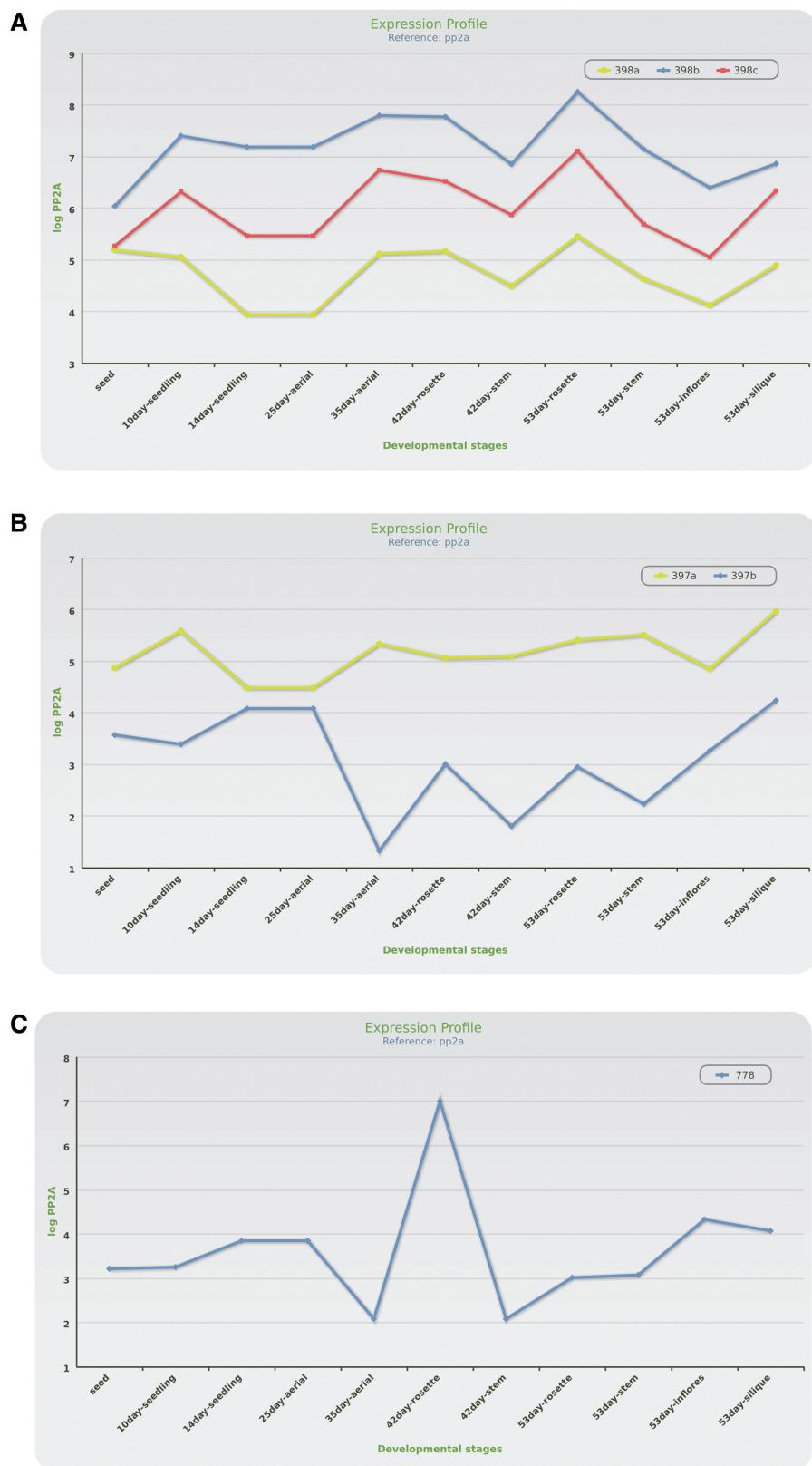


Figure 3. Examples of various microRNA gene expression profiles. The presented graphs were created with options available in mirEX interface and represent (A) miR398 and (B) miR397 gene family (see text for details). (C) Example of miR778 showing dramatic changes in expression level during growth and organ development.

the opportunity to easy access-related information from the most significant databases in the field. In every day laboratory work, this approach proved to be very efficient, and made the mirEX platform a one-stop information center for *Arabidopsis* microRNA data.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figure S1.

ACKNOWLEDGEMENTS

We thank Professor Eva Czarnecka for the helpful suggestions in writing this manuscript and Lance Verner for proofreading it.

FUNDING

The Polish Ministry of Science and Higher Education (grant 3011/B/P01/2009/37); the Faculty of Biology Adam Mickiewicz University in Poznan, Poland; the Foundation for Polish Science (FNP) within the International PhD Program co-financed from European Union Regional Development Fund (MPD 2010/3 to D.B. and J.D.). Funding for open access charge: The Faculty of Biology Adam Mickiewicz University.

Conflict of interest statement. None declared.

REFERENCES

1. Reinhart,B., Weinstein,E., Rhoades,M., Bartel,B. and Bartel,D. (2002) MicroRNAs in plants. *Genes Dev.*, **16**, 1616–1626.
2. Llave,C., Xie,Z., Kasschau,K.D. and Carrington,J.C. (2002) Cleavage of scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science*, **297**, 2053–2056.
3. Brodersen,P., Sakvarelidze-Achard,L., Bruun-Rasmussen,M., Dunoyer,P., Yamamoto,Y.Y., Sieburth,L. and Voinnet,O. (2008) Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, **320**, 1185–1190.
4. Chambers,C. and Shuai,B. (2009) Profiling microRNA expression in *Arabidopsis* pollen using microRNA array and real-time PCR. *BMC Plant Biol.*, **9**, 87.
5. Lee,H., Yoo,S.J., Lee,J.H., Kim,W., Yoo,S.K., Fitzgerald,H., Carrington,J.C. and Ahn,J.H. (2010) Genetic framework for flowering-time regulation by ambient temperature-responsive miRNAs in *Arabidopsis*. *Nucleic Acids Res.*, **38**, 3081–3093.
6. Laubinger,S., Sachsenberg,T., Zeller,G., Busch,W., Lohmann,J.U., Rätsch,G. and Weigel,D. (2008) Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **105**, 8795–8800.
7. Meyers,B.C., Tej,S.S., Vu,T.H., Haudenschild,C.D., Agrawal,V., Edberg,S.B., Ghazal,H. and Decola,S. (2004) The use of MPSS for whole-genome transcriptional analysis in *Arabidopsis*. *Genome Res.*, **14**, 1641–1653.
8. Graeber,K., Linkies,A., Wood,A.T.A. and Leubner-Metzger,G. (2011) A guideline to family-wide comparative state-of-the-art quantitative RT-PCR analysis exemplified with a Brassicaceae cross-species seed germination case study. *Plant Cell*, **23**, 2045–2063.
9. Kozomara,A. and Griffiths-Jones,S. (2010) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–D157.
10. Zhao,B., Liang,R., Ge,L., Li,W., Xiao,H., Lin,H., Ruan,K. and Jin,Y. (2007) Identification of drought-induced microRNAs in rice. *Biochem. Biophys. Res. Commun.*, **354**, 585–590.
11. Moldovan,D., Spriggs,A., Yang,J., Pogson,B.J., Dennis,E.S. and Wilson,I.W. (2010) Hypoxia-responsive microRNAs and trans-acting small interfering RNAs in *Arabidopsis*. *J. Exp. Bot.*, **61**, 165–177.
12. Zhao,M., Ding,H., Zhu,J.-K., Zhang,F. and Li,W.-X. (2011) Involvement of miR169 in the nitrogen-starvation responses in *Arabidopsis*. *New Phytol.*, **190**, 906–915.
13. Backman,T.W.H., Sullivan,C.M., Cumbie,J.S., Miller,Z.A., Chapman,E.J., Fahlgren,N., Givan,S.A., Carrington,J.C. and Kasschau,K.D. (2008) Update of ASRP: the *Arabidopsis* small RNA project database. *Nucleic Acids Res.*, **36**, D982–D985.
14. Meng,Y., Gou,L., Chen,D., Mao,C., Jin,Y., Wu,P. and Chen,M. (2010) PMiRKB: a plant microRNA knowledge base. *Nucleic Acids Res.*, **39**, D181–D187.
15. Betel,D., Wilson,M., Gabow,A., Marks,D.S. and Sander,C. (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res.*, **36**, D149–D153.
16. Cho,S., Jun,Y., Lee,S., Choi,H.-S., Jung,S., Jang,Y., Park,C., Kim,S., Lee,S. and Kim,W. (2010) miRGator v2.0: an integrated system for functional investigation of microRNAs. *Nucleic Acids Res.*, **39**, D158–D162.
17. Kaya,K.D., Karakülah,G., Yakıcıer,C.M., Acar,A.C. and Konu,O. (2011) mESAdb: microRNA expression and sequence analysis database. *Nucleic Acids Res.*, **39**, D170–D180.
18. Ritchie,W., Flamant,S. and Rasko,J.E.J. (2010) mimiRNA: a microRNA expression profiler and classification resource designed to identify functional correlations between microRNAs and their targets. *Bioinformatics*, **26**, 223–227.
19. Pant,B., Musialak-Lange,M., Nuc,P., May,P., Walther,D. and Scheible,W. (2009) Identification of nutrient-responsive *Arabidopsis* and rapeseed microRNAs by comprehensive real-time PCR profiling and small RNA sequencing. *Plant Physiol.*, **150**, 1541–1555.
20. Szarzynska,B., Sobkowiak,L., Pant,B., Balazadeh,S., Scheible,W., Mueller-Roeber,B., Jarmolowski,A. and Szweykowski-Kulinska,Z. (2009) Gene structures and processing of *Arabidopsis thaliana* HY1-dependent pri-miRNAs. *Nucleic Acids Res.*, **37**, 3083–3093.
21. Oñate-Sánchez,L. and Vicente-Carbajosa,J. (2008) DNA-free RNA isolation protocols for *Arabidopsis thaliana*, including seeds and siliques. *BMC Res. Notes*, **1**, 93.
22. Czechowski,T., Stitt,M., Altmann,T., Udvardi,M.K. and Scheible,W.-R. (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.*, **139**, 5–17.
23. Lamesch,P., Dreher,K., Swarbreck,D., Sasidharan,R., Reiser,L. and Huala,E. (2010) Using the *Arabidopsis* information resource (TAIR) to find information about *Arabidopsis* genes. *Curr. Protoc. Bioinformatics*, Chapter 1, Unit1.11.
24. Lu,Z. (2011) PubMed and beyond: a survey of web tools for searching biomedical literature. *Database*, **2011**, doi:10.1093/database/baq036.